Virus Genes
DOI 10.1007/s11262-014-1139-8

Acquisition of new protein domains by coronaviruses: analysis of overlapping genes coding for proteins N and 9b in SARS coronavirus

Aditi Shukla · Rolf Hilgenfeld

Received: 17 July 2014 / Accepted: 25 October 2014
© Springer Science+Business Media New York 2014

Abstract Acquisition of new proteins by viruses usually occurs through horizontal gene transfer or through gene duplication, but another, less common mechanism is the usage of completely or partially overlapping reading frames. A case of acquisition of a completely new protein through introduction of a start codon in an alternative reading frame is the protein encoded by open reading frame (orf) 9b of SARS coronavirus. This gene completely overlaps with the nucleocapsid (N) gene (orf9a). Our findings indicate that the orf9b gene features a discordant codon-usage pattern. We analyzed the evolution of orf9b in concert with orf9a using sequence data of betacoronavirus-lineage b and found that orf9b, which encodes the overprinting protein, evolved largely independent of the overprinted orf9a. We also examined the protein products of these genomic sequences for their structural flexibility and found that it is not necessary for a newly acquired, overlapping protein product to be intrinsically disordered, in contrast to earlier suggestions. Our findings contribute to characterizing sequence properties of newly acquired genes making use of overlapping reading frames.

Keywords SARS coronavirus · Accessory proteins · Overlapping reading frames · Orf9b · Nucleocapsid · Evolution

Introduction

A codon is composed of three nucleotides and therefore, three reading frames are, in principle, possible within the same gene. The correct reading frame is determined by the start codon, usually ATG or, in RNA genomes, AUG. Occasionally, the same stretch of genome codes for more than one protein in different reading frames; in this case, the protein products are called "overprinted" (or "ancestral") and "overprinting" (or "novel") [1]. For simplicity, proteins encoded by overlapping genes are often called "overlapping proteins" or "transframe proteins" [2], because during their translation, there occurs either a $+1$ or -1 shift in the reading frame [3]. Overlapping reading frames have been discovered in many organisms, but they are most commonly found in viruses because of their comparatively small-size genomes [3–7]. It has been proposed that translation of overlapping reading frames in RNA viruses occurs via leaky ribosomal scanning [8–13], ribosomal frameshifting [14–17], or stop-codon readthrough [18–20]. Recently, a -2 programmed ribosomal frameshift has been observed in the synthesis of an overlapping protein in members of the family *Arteriviridae* [21, 22].

In terms of evolution, overlapping genes are considered a mechanism for creating novel proteins [1, 5]. However, any point mutation occurring in an overlapping gene region affects two (or more) protein products at the same time.

Electronic supplementary material The online version of this article (doi:10.1007/s11262-014-1139-8) contains supplementary material, which is available to authorized users.

A. Shukla · R. Hilgenfeld (✉)
Institute of Biochemistry, Center for Structural and Cell Biology in Medicine, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany
e-mail: hilgenfeld@biochem.uni-luebeck.de

A. Shukla
Graduate School for Computing in Medicine & Life Sciences, University of Lübeck, Lübeck, Germany

R. Hilgenfeld
German Center for Infection Research (DZIF), University of Lübeck, Lübeck, Germany

Published online: 20 November 2014

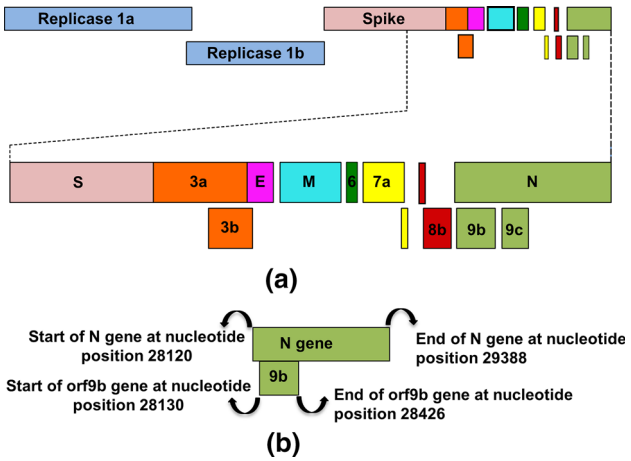

Fig. 1 **a** Schematic organization of the SARS-coronavirus genome, highlighting structural, and accessory genes. The overprinting accessory genes are indicated below their overprinted mates. **b** Overlapping N and orf9b genes of SARS-CoV with their start-stop coordinates within the genome ("end" coordinates include the stop codon)

Several studies have been performed to show the evolutionary aspects of overlapping gene sets in viruses [4, 7, 23–28]. Here, we report a case study on a set of overlapping proteins in severe acute respiratory syndrome coronavirus (SARS-CoV). SARS is characterized by acute pulmonary inflammation with a case/fatality rate of about 10 % [29, 30]. SARS-CoV first emerged in 2002 in Guangdong province, China, and developed into a widespread epidemic in the spring of 2003 [29]. It is believed that the virus originated from a bat reservoir [31–33]. Although the epidemic ended in July 2003, it is not unlikely for SARS-CoV or a similar virus to re-surface again. This view is supported by the recent emergence of a new human betacoronavirus of lineage c, Middle-East respiratory syndrome (MERS) CoV [34–37], which is transmitted to humans from dromedary camels [38–40] but may have a bat origin as well [41–43]. As of October 16, 2014, 877 laboratory-confirmed human MERS cases have been reported since September 2012, including at least 317 deaths [44].

The 30-kb single-stranded SARS-CoV RNA genome codes for at least 28 proteins. In the 3′-proximal third, small open reading frames (orfs) coding for the so-called accessory proteins are interspersed among the genes coding for the structural proteins [45, 46]. These include orf3a/b, orf6, orf7a/b, orf8a/b, orf9b, and possibly orf9c (Fig. 1) [45–49]. Accessory proteins of SARS-CoV are thought to be important players involved in viral pathogenicity [47–50]. However, reverse genetic studies have demonstrated that these proteins are not required for viral replication in cell culture or transgenic mice expressing human ACE2, the receptor for SARS-CoV [51–53]. Here we present the results of an investigation into the evolution and differential selection affecting the orf9b protein. This gene completely overlaps with the gene coding for the 422-amino-acid residue nucleocapsid (N) protein (Fig. 1). Recently, it has been proposed that leaky ribosomal scanning is responsible for the production of the overlapping orf9b protein [54]. Orf9b codes for a small protein of 98 amino-acid residues, which is found in SARS-CoV-infected cells [55]. Antibodies against this protein have been detected in the sera of SARS patients, demonstrating that the protein is produced during infection [50, 56]. There has also been a report of yet another overlapping gene, orf9c, within the nucleocapsid gene, coding for a predicted protein of 70 amino-acid residues [45, 46] (see Fig. 1). Until now, no evidence of orf9c expression has become available and since it is not well annotated, we will not include this gene in our analysis.

In case of the overlapping nucleocapsid and orf9b genes of SARS-CoV, we are in the fortunate and rare situation that a large body of sequence data is available, as many isolates of the virus and its relatives from civets and bats were analyzed during and after the outbreak of 2003 [57]. In addition, crystal structures have been determined for both the overlapping proteins [58, 59], allowing conclusions that would not be possible in most cases of overlapping viral genes. In order to shed light onto the evolution of orf9b in concert with the nucleocapsid gene, we analyzed the codon-usage patterns of the two genes and the effects of the overlap on their mutation rates as well as on the three-dimensional structures of their protein products.

Methods

Seventy full-length genomic sequences including one reference sequence of SARS-CoV (isolate Tor2) were retrieved from the GenBank database (http://www.ncbi.nlm.nih.gov/Genbank/index.html). Among these, 37 were from human SARS-CoV isolates, 15 from civet SARS-CoV, and 18 (including the newly discovered SL-CoV-WIV1 [33]) from bat betacoronaviruses lineage b. Accession numbers are given in Table S1 (Online Resource 1). Unless explicitly stated otherwise, all these viruses are collectively called "SARS-CoV" in this work. These full-length genomic sequences were parsed and corresponding gene and amino-acid sequences were collected in a local database for further analysis.

Codon-usage analysis

Codon-usage analysis was done using the "sequence analysis program" within the Sequence Manipulation Suite [60]. This program accepts one or more nucleotide

sequence(s) and returns the number and frequency of each codon type. Relative Synonymous Codon Usage (RSCU) values were calculated using the frequencies of each codon type of a nucleotide sequence. The RSCU value for a codon i is calculated by: RSCU (i) = (observed i)/(expected i), where "observed i" is the observed frequency of codon i in a gene and "expected i" is the frequency expected assuming equal usage of synonymous codons for an amino acid in a gene.

The RSCU index is a measure to assess whether a sequence shows a preference for particular synonymous codons, i.e., codons that code for the same amino acid. The comparison was done at the level of codon usage between overlapping and non-overlapping coding regions. The RSCU values obtained from the overlapping orf9a/9b genes were then compared with those of the non-overlapping regions of the same viral genome by means of the Pearson correlation coefficient "r" [61]. Values of r range from −1 to +1, reflecting a completely different and identical degree in the usage of synonymous codons, respectively. Discordant usage suggests the gene to be relatively new [62, 63].

Mutation rate analysis

Mutation rate analysis was performed by first aligning both the nucleotide and protein sequences using ClustalX version 2.1 [64]. Redundant sequences (from multiple human patients) were manually removed. DnaSP 5.10 [65] was used to calculate the number of synonymous and nonsynonymous substitutions at the overlapping gene regions of the N gene and the orf9b gene.

Entropy-plotting of alignments

Entropy-plotting of alignments was done to determine the variations occurring in the overprinted N protein and the overprinting orf9b protein. Variation in the three sets of nucleotide sites of a codon and the variation in the corresponding amino-acid residues in the overlapping proteins were studied by plotting the entropy (variability) of the aligned overlapping nucleotide and protein sequences, as implemented in the BioEdit software v7.0.0. [66]. The entropy is defined as a measure of uncertainty at each position in a set of aligned nucleotide or protein sequences [66]. The cumulative entropy is the sum of all the entropy values calculated at each position in a sequence. For calculating the entropy, sequences are treated as a matrix of characters and the maximum number of different characters found in a column (column of aligned sets of nucleotide or protein sequences) defines the maximum total uncertainty or the "entropy" [66]. The entropy H is calculated by $H(l) = -\sum f(b,l)\ln(f(b,l))$, where H(l) is the uncertainty (entropy) at position l, b represents a residue (out of the allowed choices for the sequence under investigation), and f(b,l) is the frequency at which residue b is found at position l. We determined the frequency of substitutions at each codon site in the overlapping region of the nucleocapsid/orf9b gene to find out the evolutionary strategy followed by this set of overlapping genes in SARS-CoV. There are 98 codons in the overlapping region and we studied the variation of nucleotides (294 codon positions) and their corresponding amino acids in the overlapping nucleocapsid and orf9b gene region of 70 SARS-CoV genomes.

Inspection of crystal structures and prediction of disorder

Inspection and presentation of crystal structures of the N-terminal domain (NTD) of the N protein (orf9a) [58] and the orf9b protein of SARS-CoV [59] were performed using the program Pymol (Schrödinger), in order to assess possible consequences of the overlapping genes for the three-dimensional structures of the protein products. Disorder prediction was also carried out for these overlapping proteins because crystallographic information is not available for the N-terminal 46 residues of the overprinted N protein. For this purpose, the DisProt VSL2 intrinsic-disorder prediction program was used [67].

Results

Codon usage in the overlapping gene set

The first step of our analysis was to establish a relationship between the codon usage in overlapping and non-overlapping genes in the betacoronavirus SARS-CoV. Two-thirds of the SARS-CoV genome comprise orf1ab, which encodes the viral polyproteins pp1a and pp1ab. The 3′-proximal third comprises orfs encoding the structural proteins, i.e., spike (S), envelope (E), membrane (M), and nucleocapsid (N), as well as several accessory proteins as described previously [45, 46]. The non-overlapping regions of the genome (Fig. 1) were combined into a single unit including the orf1a, orf1b, spike, envelope, membrane, and orf6 genes. On the other hand, the genes under study here, full-length nucleocapsid and the overlapping orf9b, were considered distinct sets of data. The remaining accessory proteins were not included in this comparison because they contain partially overlapping regions.

The subsequent correlation analysis (Table 1) shows that the internal overlapping gene (orf9b) exhibits a choice of synonymous codons highly different from that occurring in the non-overlapping gene set of SARS-CoV, exhibiting an r value of −0.01. For example, of the eight proline residues in the orf9b protein, five (63 %) are coded by

Table 1 Correlation between the codon-usage patterns of the overlapping N and its overprinting internal genes of SARS-CoV, as well as of other members of the genus *Betacoronavirus,* i.e., BCoV, MHV, and MERS-CoV, each with the non-overlapping coding regions in their genome

Betacoronavirus	Proteins	Number of amino-acid residues	Correlation coefficient (r)
SARS-CoV	Nucleocapsid	422	0.62
	Orf9b	98	−0.01
BCoV	Nucleocapsid	448	0.67
	internal protein	207	−0.11
MHV	Nucleocapsid	455	0.66
	internal protein	136	0.00
MERS-CoV	Nucleocapsid	411	0.58
	hypothetical internal protein	112	−0.13

Table 2 Synonymous and nonsynonymous substitutions in overlapping and non-overlapping regions of the SARS-CoV nucleocapsid gene and in the orf9b gene

Gene regions of:	K_a	K_s	$K_a/K_s = \omega$
Nucleocapsid (overlapping part)	0.41	0.73	0.56
Nucleocapsid (non-overlapping part)	0.37	0.59	0.62
Orf9b	0.53	0.43	1.23

CCC, whereas in the non-overlapping gene set, less than 9 % of proline residues use this codon. A much higher degree of concordant relationship is seen between the overprinted N gene of SARS-CoV and the non-overlapping gene set (*r* value of 0.62).

Codon-usage analysis for other betacoronaviruses containing a hypothetical overlapping "internal" gene within their nucleocapsid gene was also carried out for comparison. The nucleocapsid genes of Bovine Coronavirus (BCoV, NCBI accession number NC_003045), Mouse Hepatitis Virus (MHV, AC_000192), and Middle-East Respiratory Syndrome coronavirus (MERS-CoV, NC_019843) display a similar positive correlation when compared to the rest of the genome, with *r* values of 0.67, 0.66, and 0.57, respectively, whereas their corresponding "internal" genes have *r* values of −0.11, 0.00, and −0.13, respectively (Table 1).

Effect of the overlap on the mutation rate in the N and orf9b genes and evolutionary strategy adopted by this overlapping gene set

The ratio ω of nonsynonymous (K_a) to synonymous (K_s) nucleotide substitution rates is an indicator of selective pressure on genes. A ratio significantly greater than 1 indicates positive selective pressure. A ratio around 1 indicates either neutral evolution at the protein level or an averaging of sites under positive and negative selective pressure. A ratio less than 1 indicates pressure to conserve protein sequence, i.e., "purifying selection" [68].

The overlapping gene regions of nucleocapsid and orf9b show differences in the evolutionary rates. The overprinted region of the nucleocapsid gene in SARS-CoV has a K_a/K_s ratio of 0.56 (the K_a/K_s ratio is 0.62 for the non-overlapping part of the N gene). On the other hand, the orf9b gene has a K_a/K_s ratio greater than 1 (Table 2), which means that the orf9b protein is subject to positive selection pressure and is evolving at a faster rate, compared to the overprinted N gene. Remarkably, due to the difference in frame phase (see below), the same stretch of genome has thus different evolutionary rates when coding for each of the two different proteins. This observation prompted us to analyze the nucleotide variations at each of the three nucleotide positions of the codons.

Upon a point mutation, the position at which nucleotide substitution occurs within a codon reflects whether the substitution would be synonymous or not. When there is a nucleotide substitution at the first codon position, it causes an amino-acid change in 60 out of 64 cases. The four exceptions occur because of the partial degeneration of codons at this position. At this codon site, there are four possibilities, which would result in a synonymous substitution: two each in codons for leucine (UUA vs. CUA, UUG vs. CUG) and arginine (AGA vs. CGA, AGG vs. CGG). When there is a nucleotide substitution at the second codon position, it results in an amino-acid change in 63 out of 64 cases. At this codon site, the only substitution that is synonymous occurs in the stop codons UAA versus UGA. Lastly, a nucleotide substitution occurring at the third codon position causes a change in amino acid in only 16 out of 64 cases because of codon degeneracy at the third nucleotide position [25, 62, 63].

As a result of leaky ribosomal scanning [54], translation of orf9b begins at the 10th nucleotide position of the N gene (Fig. 2a), resulting in a +1 difference in reading frame between orf9b ((+1)-phase frame) and the N gene (0-phase frame). Thus, the first nucleotide position of an N codon corresponds to the third nucleotide position of an

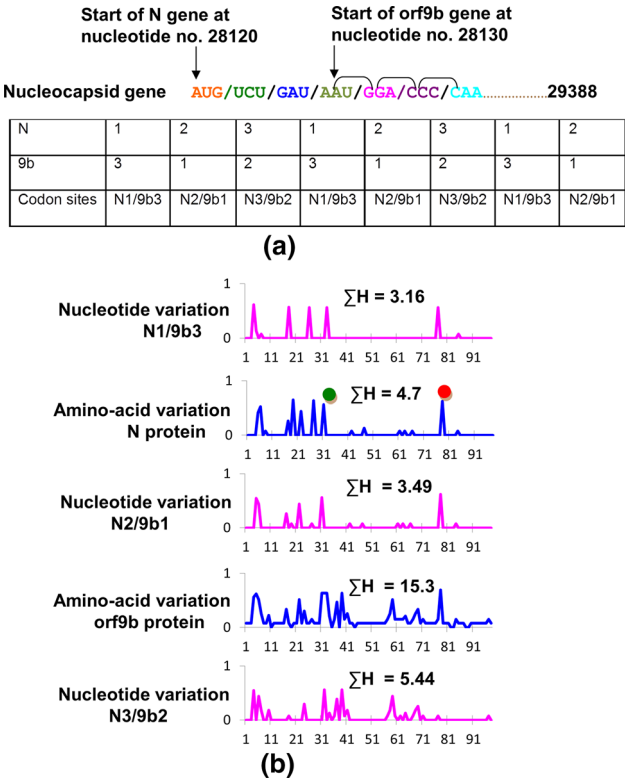

Fig. 2 **a** *Top*: The 5′ ends of the SARS-CoV N and orf9b genes. At nucleotide no. 10 of the N gene, translation of the overlapping orf9b gene begins, resulting in a phase difference of +1 for this gene relative to the N gene. *Bottom*: Codon-site substitutions in the two genes. Three types of substitution have to be distinguished: N2/9b1, N3/9b2, N1/9b3. **b** Variation of three sets of nucleotides (in *magenta*): N1/9b3, N2/9b1, and N3/9b2, in relation to the amino-acid variations (in *blue*) in the overlapping nucleocapsid and orf9b proteins. The *x*-axis represents the codon sites in case of graphs 1, 3, and 5, i.e., nucleotide variations, whereas in case of graphs 2 and 4, the *x*-axis represents the amino-acid residue number. Note that the N protein overlaps with orf9b between its residues 4 and 101; however, in graph 2, which represents the amino-acid variations in the N protein, the *x*-axis is calibrated from 1 to 98 in order to facilitate the comparison with orf9b. The *y*-axis represents entropy. The *green dot* indicates the one case of synonymous N1/9b3 substitution that does not lead to an amino-acid exchange in the N protein because of the partial degeneration of the first nucleotide position in a codon (AGA and CGA both code for Arg). The *red dot* indicates a case of a two-nucleotide difference as a result of an N1/9b3 and an N2/9b1 substitution that leads to an amino-acid exchange in the N protein. All bat betacoronaviruses of lineage b (with the exception of SL-CoV WIV1 [33]) have Lys at this position, whereas all civet and human SARS-CoV isolates as well as bat SL-CoV WIV1 have Pro (see text)

orf9b codon (N1/9b3), the second position in an N codon corresponds to the first nucleotide position in an orf9b codon (N2/9b1), and the third position in an N codon corresponds to the second nucleotide position in an orf9b codon (N3/9b2) (Fig. 2a). Hence, as described above, a nucleotide substitution in the first position of a nucleocapsid codon (N1/9b3) is likely to cause an amino-acid change in N, but not in orf9b, whereas substitutions in the third position of an N codon (N3/9b2) are probably nonsynonymous in orf9b, but synonymous in N.

The nucleotide variation at the 98 N1/9b3, N2/9b1, and N3/9b2 positions (see Fig. 2a) of the overlapping N/orf9b gene region for all 70 sequences is shown in Fig. 2b. We find that the value of cumulative mutational frequency ($\sum$(H); see "Methods" section) of the overlapping region of the nucleocapsid protein is 4.7 and that of the orf9b protein is 15.3. This difference in the frequency of mutation was somewhat expected, based on the different ω values for the two genes (see Table 2). Moving on to the nucleotide level, we obtain cumulative entropy values of 3.16, 3.49, and 5.44 for the N1/9b3, N2/9b1, and N3/9b2 codon positions, respectively. The graphs are calibrated in the range of 0–1 for accurate comparison of the results of protein sequences with nucleotide sequences (Fig. 2b).

The higher rate of amino-acid variation in the orf9b protein is largely determined by nucleotide substitutions at the N3/9b2 sites. All the N3/9b2 nucleotide variations translate to amino-acid changes in the orf9b protein but are silent in the nucleocapsid protein. Amino-acid variations in the nucleocapsid protein are determined by substitutions of N1/9b3 nucleotides. In one instance, an N1/9b3 nucleotide variation results in a synonymous mutation in the nucleocapsid protein (Fig. 2b, green dot). The amino acid at this position is arginine and this phenomenon occurs due to partial degeneration of the first nucleotide position as explained above.

There are a few N2/9b1 mutations that impose amino-acid variations in both the proteins. An interesting variation, corresponding to a concomitant N1/9b3 and N2/9b1 exchange, results in an amino-acid difference at position 81 of the nucleocapsid protein, within its well-ordered and overprinted part. All known genomic sequences of bat betacoronaviruses of lineage b feature the AAA triplet (coding for Lys) here, whereas all isolates of civet and human SARS-CoV have CCA (coding for Pro). The exception among the bat beta-CoVs of lineage b is the newly discovered SL-CoV WIV1, which is proposed to be the likely originator of SARS-CoV [33]; the N gene of this virus also has CCA coding for Pro at this position. Thus, there is a two-nucleotide difference between the codons in the bat CoVs (except SL-CoV WIV1) on the one hand and human or civet SARS-CoV on the other (see red dot in Fig. 4). In the orf9b protein, the corresponding codon (shifted by +1 in frame) is AAG (coding for Lys) in the bat betacoronaviruses of lineage b, and CAG (coding for Gln) in the civet and human SARS-CoV sequences as well as in SL-CoV WIV1 [33]. Thus, only the N2/9b1 variation results in an amino-acid change in the orf9b protein and the N1/9b3 nucleotide variation is silent.

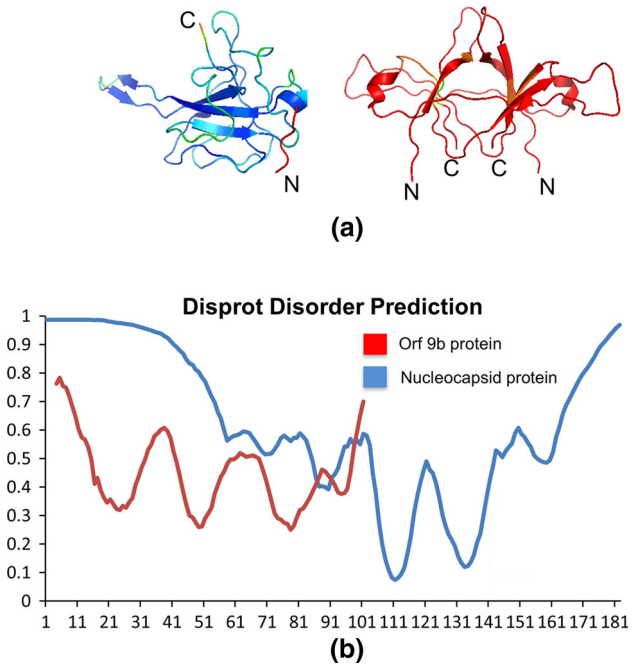

Fig. 3 **a** Structures of the SARS-CoV proteins investigated in this study, colored according to the B-factor (color scheme used is VIBGYOR, where *Violet* depicts the minimum and *red* depicts the maximum value of B-factor) averaged for each amino-acid residue; (*left*) NTD of nucleocapsid protein (overall average B-factor 11.19 $Å^{2;}$ PDB code 2OFZ [58]); (*right*) Dimer of the orf9b protein (overall average B-factor 100.8 $Å^2$, PDB code 2CME [59]). **b** Disorder prediction result for the NTD of the SARS-CoV nucleocapsid protein and its overprinting counterpart, the orf9b protein, calculated using the program DisProt VSL2B [67]. The degree of order–disorder lies within the range of 0 (well ordered) to 1 (highly disordered)

Effect of overlaps on the three-dimensional structures of the protein products

Crystal structures are available for both the N-terminal RNA-binding domain (NTD) of the SARS-CoV nucleocapsid protein [58] (PDB code: 2OFZ) and the orf9b protein [59] (PDB code: 2CME). The N-terminal 46 residues of the NTD have been excluded from the fragment that was crystallized, as they are believed to be disordered on the basis of secondary structure prediction, limited proteolysis experiments, and sequence conservation. The well-ordered global NTD (residues 47–175) comprises an antiparallel β-sheet core and a β-hairpin protruding from it (Fig. 3a). The orf9b protein forms a two-fold symmetric dimer comprising two adjacent β-sheets (Fig. 3a). In the central hydrohobic cavity between the monomers, electron density for a lipid molecule was detected [59]. The global part of the nucleocapsid NTD exhibits reduced flexibility, as evident from the atomic temperature factors (B-factors) for the polypeptide. The average B-factor for the NTD (residues 47–175) is 11.2 $Å^2$ [58]. In contrast, the orf9b protein appears to be much more flexible; its average B-factor is 100.8 $Å^2$ [59]. Even though a number of factors contribute to the B value, in particular the degree of disorder of the crystal, rigid-body movements of the molecules, experimental errors, etc., it is evident from this huge difference between the B values for the two structures that the orf9b polypeptide chain is very flexible. In line with this, the segments between residues 1–8 and 26–37 in the orf9b protein are not visible in the electron-density maps. This result supports the notion that overprinting proteins tend to show higher flexibility and disorder [1]. In contrast, all residues are well defined by electron density in the NTD crystal structure of the overlapping N protein. However, it should also be mentioned that probably, many residues are disordered or highly flexible among the N-terminal 46 residues of the NTD, which have not been included in the construct employed in crystal structure determination but belong to the overlapping region (with the exception of the first three residues) as seen in a prediction using the program DisProt VSL2B [67] (Fig. 3b).

We also made an attempt to locate the amino-acid variations, which we identified in the 70 sequences of our data set, in the three-dimensional structures of the nucleocapsid [58] and orf9b [59] proteins. In the overprinted part of the nucleocapsid NTD, the majority of mutations occur in the unstructured N-terminal region (residues 1–46). Residue 81, which is Lys in all sequenced bat betacoronaviruses of lineage b (except SL-CoV WIV1) but Pro in all civet and human SARS-CoV strains, is located at position 2 of a surface-exposed type-III β-turn of the sequence Gly-Pro-Asp-Asp. In the orf9b structure, the corresponding residue (Gln in SARS-CoV and SL-CoV WIV1) is not defined by electron density and hence part of a presumably disordered region. In fact, in the orf9b protein, all mutations occur in the regions that have been reported to be disordered [59]. Thus, the general observation that mutations are more commonly localized in regions of no regular secondary structure (such as loops etc.) rather than in α-helices or β-strands, remains valid for this overprinting protein.

Discussion

Previous work has suggested codon usage as a measure to determine the relative age of a gene. A discordant relationship in the codon usage of a particular gene, when compared with the rest of the genes, suggests that the gene has evolved recently [1, 25, 28, 62, 63]. It has been shown that this phenomenon can be applied to identify overprinting viral proteins [63]. Research on overlapping protein products in RNA viruses has suggested that proteins of the overprinting genes have a tendency to be structurally

disordered [1]. Taking into account these observations, we have analyzed the sequence and structural properties of the overlapping proteins nucleocapsid and orf9b in SARS-CoV, in the hope of gaining insight into the creation of novel proteins in RNA viruses. Studies revealing differential selective pressure during the evolution of overlapping viral genes [23–27, 62] led us to probe the selection pressure acting on the overlapping N and orf9b genes.

Codon-usage analysis shows that the overprinting orf9b has a discordant codon-usage pattern whereas the overprinted region of the nucleocapsid gene has a codon-usage pattern similar to the non-overlapping regions of the SARS-CoV genome. The discordant codon-usage pattern suggests a relatively more recent acquisition of the orf9b gene [62, 63]. Moreover, no sequence similarity exists between the orf9b protein and any other known proteins [45, 46]. Among other coronaviruses, internal open reading frames within the nucleocapsid gene have been reported for members of *Betacoronavirus* lineage a, for example Bovine Coronavirus, Mouse Hepatitis Virus, and human coronaviruses HKU1 and OC43 [69–71]. Moreover, they can also be found in members of *Betacoronavirus* lineage c, i.e., MERS Coronavirus [36] as well as bat coronaviruses HKU4 and HKU5. However, there is little in common between the orf9b gene of SARS-CoV and these so-called "internal" genes [49]. Hence, we can conclude that orf9b is a novel gene.

A previous study performed on the translation mechanism employed in orf9b expression describes the presence of a sub-optimal Kozak sequence for the N gene [54]. In contrast, the start codon of the orf9b gene resides within an optimal Kozak sequence, but the properties of the second initiation site in leaky ribosomal scanning do not influence the decision of the ribosome to stop at or bypass the first AUG [9]. Xu et al. [54] showed that the expression level of the orf9b gene is relatively weaker when compared to that of the N gene. This is probably caused by only a fraction of ribosomes bypassing the first AUG, but may also be influenced by the fact that the codon usage of orf9b is different from that of the rest of the SARS-CoV genes (Table 1). In addition, although orf9b production appears to depend on leaky ribosomal scanning, we cannot exclude that the expression of this gene is modulated by other SARS-CoV proteins, as described very recently for members of the family *Arteriviridae* [21, 22]. The orf9b protein has been shown to interact with several other SARS-CoV proteins, for example Nsp5, Nsp14, and the orf6 protein [52, 72]. It remains to be investigated whether any of these can transactivate or suppress the expression of the alternating gene, orf9b.

The evolution of overlapping, frame-shifted genes is subject to extra constraints. A slower rate of evolution has been demonstrated in overlapping genes in a number of viruses [7, 23–26]. There is an imposing constraint in evolution of overlapping genes due to the fact that a favorable or even neutral substitution in one reading frame could prove harmful for the other reading frame. Therefore, even a synonymous, favorable, or neutral nucleotide substitution in one reading frame might be discarded, as it could be deleterious in the other reading frame. As a result, positive selection of overlapping genes in general is severely restricted [7, 23–26]. However, here we demonstrated that the overlapping region of the N gene is rather evolutionary conserved when compared to the orf9b gene. Orf9b features a higher evolutionary rate that is attained mainly via N3/9b2 substitutions. This mechanism of independent evolution is similar to the mechanism of "independent adaptive selection" recently described for the gene encoding the Hepatitis B Virus (HBV) surface protein which completely overlaps with the polymerase gene [27].

A structural comparison of the overlapping NTD of the nucleocapsid protein with the orf9b protein indicates the latter to be more flexible (Fig. 3). However, 50 % of the overlapping orf9b protein corresponds to the structurally undetermined segment 1–46 of the NTD of the nucleocapsid protein. This region is predicted to be intrinsically disordered (see Fig. 3b) [73]. Consequently, the overprinting orf9b protein is more ordered in its N-terminal half than the overprinted N-terminal segment 1–46 of the nucleocapsid NTD. More studies on overlapping viral proteins with known three-dimensional structures would be necessary to refine our understanding of the evolutionary and structural constraints caused by this phenomenon.

Currently, the function of the orf9b protein is unknown, although its crystal structure [59] suggests that it may bind lipids. Since the protein evolved with a high rate and independently of the overprinted nucleocapsid protein, it may have the potential to acquire new functions in a relatively short time. The analysis presented here may form a basis to follow the future evolution of orf9b and a starting point for investigating the so-called "internal genes" overlapping with the nucleocapsid gene in other coronaviruses, including MERS-CoV. Beyond its importance for understanding coronavirus evolution, our study may have implications for the analysis of overlapping reading frames in other viruses as well.

Acknowledgments We thank Dr. Zhengli Shi of the Wuhan Institute of Virology, Chinese Academy of Sciences, for sharing the sequence of the SL-CoV WIV1 nucleocapsid gene with us prior to publication. This project has been supported by the "Graduate School for Computing in Medicine & Life Sciences" funded by Germany's Excellence Initiative [DFG GSC 235/2].

References

1. C. Rancurrel, M. Khosravi, A.K. Dunker, P. Romero, D. Karlin, Overlapping genes produce proteins with unusual sequence properties and offer Insight into de novo protein creation. J. Virol. **83**, 10719–10736 (2009)
2. E.P. Plant, Ribosomal frameshift signals in viral genomes, *in Viral Genomes—Molecular Structure, Diversity, Gene Expression Mechanisms and Host-Virus Interactions*, ed. by M. Garcia (ISBN, 978-953-51-0098-0, InTech, 2012). doi:10.5772/26550. http://www.intechopen.com/books/viral-genomes-molecular-structure-diversity-gene-expression-mechanisms-and-host-virus-interactions/frameshift-signals-in-viral-genomes. Accessed 16 July 2014. Accessed 16 July 2014
3. R. Belshaw, O.G. Pybus, A. Rambaut, The evolution of genome compression and genomic novelty in RNA viruses. Genome Res. **10**, 1496–1504 (2007)
4. K.I. Jordan, B.A. Sutter, M.A. McClure, Molecular evolution of the paramyxoviridae and the rhabdoviridae multiple-protein-encoding P gene. Mol. Biol. Evol. **17**, 75–86 (2000)
5. D.C. Krakauer, Stability and evolution of overlapping genes. Evolution **54**, 731–739 (2000)
6. N. Chirico, A. Vianelli, R. Belshaw, Why genes overlap in viruses? Proc. Biol. Sci. **277**, 3809–3817 (2010)
7. T. Miyata, T. Yasunaga, Evolution of overlapping genes. Nature **272**, 532–535 (1978)
8. M. Kozak, The scanning model for translation, an update. J. Cell Biol. **108**, 229–241 (1989)
9. M. Kozak, Pushing the limits of the scanning mechanism for initiation of translation. Gene **299**, 1–34 (2002)
10. M. Kozak, Initiation of translation in prokaryotes and eukaryotes. Gene **234**, 187–208 (1999)
11. L.A. Ryabova, M.M. Pooggin, T. Hohn, Translation reinitiation and leaky scanning in plant viruses. Virus Res. **119**, 52–62 (2006)
12. S. Zou, E.G. Brown, Translation of the reovirus M1 gene initiates from the first AUG codon in both infected and transfected cells. Virus Res. **40**, 75–89 (1996)
13. D. Matsuda, T.W. Dreher, Close spacing of AUG initiation codons confers dicistronic character on a eukaryotic mRNA. RNA **12**, 1338–1349 (2006)
14. T. Jacks, H.D. Madhani, F.R. Masiarz, H.E. Varmus, Signals for ribosomal frameshifting in Rous Sarcoma virus gag-pol region. Cell **55**, 447–458 (1988)
15. I. Brierley, P. Digard, S. Inglis, Characterization of an efficient ribosomal frameshifting signal, requirement for an RNA pseudoknot. Cell **57**, 537–547 (1989)
16. I. Brierely, F.J. Dos Ramos, Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. Virus Res. **119**, 29–42 (2006)
17. J. Dinman, Mechanisms and implications of programmed translational frameshifting. Wiley Interdiscip. Rev. RNA **3**, 661–673 (2012). doi:10.1002/wrna.1126
18. A. Honigman, *cis* acting RNA sequences control the gag-pol translation readthrough in murine leukemia virus. Virology **183**, 313–319 (1991)
19. M. Orlova, Reverse transcriptase of moloney murine leukemia virus binds to eukaryotic release factor 1 to modulate suppression of translational termination. Cell **115**, 319–331 (2003)
20. H. Beier, UAG readthrough during TMV RNA translation, isolation and sequence of two tRNAs with suppressor activity from tobacco plants. EMBO J. **3**, 351–356 (1984)
21. Y. Fang, E.E. Treffers, Y. Li, A. Tas, Z. Sun, Y. van der Meer, A.H. de Ru, P.A. van Veelen, J.F. Atkins, E.J. Snijder, A.E. Firth, Efficient -2 frameshifting by mammalian ribosomes to synthesize an additional arterivirus protein. Proc. Natl. Acad. Sci. USA. **109**, E2920–E2928 (2012). doi:10.1073/pnas.1211145109
22. Y. Li, E.E. Treffers, S. Napthine, A. Tas, L. Zhu, Z. Sun, S. Bell, B.L. Mark, P.A. van Veelen, M.J. van Hemert, A.E. Firth, I. Brierley, E.J. Snijder, Y. Fang, Transactivation of programmed ribosomal frameshifting by a viral protein. Proc. Natl. Acad. Sci. USA. **111**, E2172–E2181 (2014). doi:10.1073/pnas.1321930111
23. A. Pavesi, Detection of signature sequences in overlapping genes and prediction of a novel overlapping gene in hepatitis G virus. J. Mol. Evol. **50**, 284–295 (2000)
24. Y. Fujii, K. Kiyotani, T. Yoshida, T. Sakaguchi, Conserved and non-conserved regions in the Sendai virus genome, evolution of a gene possessing overlapping reading frames. Virus Genes **22**, 47–52 (2001)
25. A. Pavesi, Origin and evolution of overlapping genes in the family *Microviridae*. J. Gen. Virol. **87**, 1013–1017 (2006)
26. M. Mizokami, E. Orito, K. Ohba, K. Ikeo, J.Y. Lau, T. Gojobori, Constrained evolution with respect to gene overlap of hepatitis B virus. J. Mol. Evol. **44**(Suppl 1), S83–S90 (1997)
27. H.L. Zaaijer, F.J. van Hemert, M.H. Koppelman, V.V. Lukashov, Independent evolution of overlapping polymerase and surface protein genes of hepatitis B virus. J. Gen. Virol. **88**, 2137–2143 (2007)
28. N. Sabath, A. Wagner, D. Karlin, Evolution of viral proteins originated *de novo* by overprinting. Mol. Biol. Evol. **29**, 3767–3780 (2012)
29. V.C.C. Cheng, J.F.W. Chan, K.K.W. To, K.Y. Yuen, Clinical management and infection control of SARS. Antiviral Res. **100**, 407–419 (2013)
30. R. Hilgenfeld, J.S.M. Peiris, From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. Antiviral Res. **100**, 286–295 (2013)
31. W. Li, Z. Shi, M. Yu, W. Ren, C. Smith, J.H. Epstein, H. Wang, G. Crameri, Z. Hu, H. Zhang, J. Zhang, J. McEachern, H. Field, P. Daszak, B.T. Eaton, S. Zhang, L.F. Wang, Bats are natural reservoirs of SARS-like coronaviruses. Science **310**, 676–679 (2005)
32. J.F. Drexler, V.M. Corman, C. Drosten, Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. Antiviral Res. **101**, 45–56 (2014)
33. X.Y. Ge, J.L. Li, X.L. Yang, A.A. Chmura, G. Zhu, J.H. Epstein, J.K. Mazet, B. Hu, W. Zhang, C. Peng, Y.J. Zhang, C.M. Luo, B. Tan, N. Wang, Y. Zhu, G. Crameri, S.Y. Zhang, L.F. Wang, P. Daszak, Z.L. Shi, Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature **503**, 535–538 (2013)
34. A.M. Zaki, S. van Boheemen, T.M. Bestebroer, A.D.M.E. Osterhaus, R.A.M. Fouchier, Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N. Engl. J. Med. **367**, 1814–1820 (2012)
35. R.J. de Groot, S.C. Baker, R.S. Baric, C.S. Brown, C. Drosten, L. Enjuanes, R.A. Fouchier, M. Galiano, A.E. Gorbalenya, Z.A. Memish, S. Perlman, L.L. Poon, E.J. Snijder, G.M. Stephens, P.C. Woo, A.M. Zaki, M. Zambon, J. Ziebuhr, Middle East Respiratory Syndrome coronavirus (MERS-CoV); Announcement of the coronavirus study group. J. Virol. **87**, 7790–7792 (2013)
36. S. van Boheemen, M. de Graaf, C. Lauber, T.M. Bestebroer, V.S. Raj, A.M. Zaki, A.D.M.E. Osterhaus, B.L. Haagmans, A.E. Gorbalenya, E.J. Snijder, R.A.M. Fouchier, Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. MBio **3**, e00473–e00512 (2012) doi:10.1128/mBio.00473-12
37. M. Cotten, T.T. Lam, S.J. Watson, A.L. Palser, V. Petrova, P. Grant, O.G. Pybus, A. Rambaut, Y. Guan, D. Pillay, P. Kellam, E. Nastouli, Full-genome deep sequencing and phylogenetic analysis of novel human betacoronavirus. Emerg. Infect. Dis. **19**, 736–742 (2013)
38. C.B. Reusken, B.L. Haagmans, M.A. Müller, C. Gutierrez, G.J. Godeke, B. Meyer, D. Muth, V.S. Raj, L. Smits-De Vries, V.M. Corman, J.F. Drexler, S.L. Smits, Y.E. El Tahir, R. De Sousa, J.

van Beek, N. Nowotny, K. van Maanen, E. Hidalgo-Hermoso, B.J. Bosch, P. Rottier, A. Osterhaus, C. Gortázar-Schmidt, C. Drosten, M.P. Koopmans, Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. Lancet Infect. Dis. **13**, 859–866 (2013)

39. B. Meyer, M.A. Müller, V.M. Corman, C.B. Reusken, D. Ritz, G.J. Godecke, E. Lattwein, S. Kallies, A. Simens, J. van Beek, J.F. Drexler, D. Muth, B.J. Bosch, U. Wernery, M.P. Koopmans, R. Wernery, C. Drosten, Antibodies against MERS coronavirus in dromedary camels, United Arab Emirates, 2003 and 2013. Emerg. Infect. Dis. **20**, 552–559 (2014)
40. B.L. Haagmans, S.H. Al Dhahiry, C.B. Reusken, V.S. Raj, M. Galiano, R. Myers, G.J. Godeke, M. Jonges, E. Farag, A. Diab, H. Ghobashy, F. Alhajri, M. Al-Thani, S.A. Al-Marri, H.E. Al Romaihi, A. Al Khal, A. Bermingham, A.D. Osterhaus, M.M. Al-Hajri, M.P. Koopmans, Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. Lancet Infect. Dis. **14**, 140–145 (2014)
41. A. Annan, H.J. Baldwin, V.M. Corman, S.M. Klose, M. Owusu, E.E. Nkrumah, E.K. Badu, P. Anti, O. Agbenyega, B. Meyer, S. Oppong, Y.A. Sarkodie, E.K. Kalko, P.H. Lina, E.V. Godlevska, C. Reusken, A. Seebens, F. Gloza-Rausch, P. Vallo, M. Tschapka, C. Drosten, J.F. Drexler, Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe. Emerg. Infect. Dis. **19**, 456–459 (2013)
42. N.L. Ithete, S. Stoffberg, V.M. Corman, V.M. Cottontail, L.R. Richards, M.C. Schoeman, C. Drosten, J.F. Drexler, W. Preiser, Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. Emerg. Infect. Dis. **19**, 1697–1699 (2013). doi:10.3201/eid1910.130946
43. Y. Yang, L. Du, C. Liu, L. Wang, C. Ma, J. Tang, R.S. Baric, S. Jiang, F. Li, Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. Proc. Natl. Acad. Sci. USA. **111**, 12516–12521 (2014). doi:10.1073/pnas.1405889111
44. World Health Organization, Global Alert and Response (GAR). Middle East respiratory syndrome coronavirus (MERS-CoV)—summary updates, http://www.who.int/csr/don/2014_07_23_mers/en/. Accessed 09 Sep 2014
45. M.A. Marra, S.J. Jones, C.R. Astell, R.A. Holt, A. Brooks-Wilson, Y.S. Butterfield, J. Khattra, J.K. Asano, S.A. Barber, S.Y. Chan, A. Cloutier, S.M. Coughlin, D. Freeman, N. Girn, O.L. Griffith, S.R. Leach, M. Mayo, H. McDonald, S.B. Montgomery, P.K. Pandoh, A.S. Petrescu, A.G. Robertson, J.E. Schein, A. Siddiqui, D.E. Smailus, J.M. Stott, G.S. Yang, F. Plummer, A. Andonov, H. Artsob, N. Bastien, K. Bernard, T.F. Booth, D. Bowness, M. Czub, M. Drebot, L. Fernando, R. Flick, M. Garbutt, M. Gray, A. Grolla, S. Jones, H. Feldmann, A. Meyers, A. Kabani, Y. Li, S. Normand, U. Stroher, G.A. Tipples, S. Tyler, R. Vogrig, D. Ward, B. Watson, R.C. Brunham, M. Krajden, M. Petric, D.M. Skowronski, C. Upton, R.L. Roper, The genome sequence of the SARS-associated coronavirus. Science **300**, 1399–1404 (2003)
46. P.A. Rota, M.S. Oberste, S.S. Monroe, W.A. Nix, R. Campagnoli, J.P. Icenogle, S. Peñaranda, B. Bankamp, K. Maher, M.H. Chen, S. Tong, A. Tamin, L. Lowe, M. Frace, J.L. DeRisi, Q. Chen, D. Wang, D.D. Erdman, T.C. Peret, C. Burns, T.G. Ksiazek, P.E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Günther, A.D. Osterhaus, C. Drosten, M.A. Pallansch, L.J. Anderson, W.J. Bellini, Characterization of a novel coronavirus associated with Severe Acute Respiratory Syndrome. Science **300**, 1394–1399 (2003)
47. K. Narayanan, C. Huang, S. Makino, SARS coronavirus accessory proteins. Virus Res. **133**, 113–121 (2008)
48. R. McBride, B.C. Fielding, The role of severe acute respiratory syndrome (SARS)-coronavirus accessory proteins in virus pathogenesis. Viruses **4**, 2902–2923 (2012)
49. D.X. Liu, T.S. Fung, K.K. Chong, A. Shukla, R. Hilgenfeld, Accessory proteins of SARS-CoV and other coronaviruses. Antiviral Res. **109**, 97–109 (2014)
50. Y.J. Tan, S.G. Lim, W. Hong, Understanding the accessory viral proteins unique to the severe acute respiratory syndrome (SARS) coronavirus. Antiviral Res. **72**, 78–88 (2006)
51. B. Yount, R.S. Roberts, A.C. Sims, D. Deming, M.B. Frieman, J. Sparks, M.R. Denison, N. Davis, R.S. Baric, Severe acute respiratory syndrome coronavirus group-specific open reading frames encode nonessential functions for replication in cell cultures and mice. J. Virol. **79**, 14909–14922 (2005)
52. M.L. Dediego, L. Pewe, E. Alvarez, M.T. Rejas, S. Perlman, L. Enjuanes, Pathogenicity of severe acute respiratory coronavirus deletion mutants in hACE-2 transgenic mice. Virology **376**, 379–389 (2008)
53. A. von Brunn, C. Teepe, J.C. Simpson, R. Pepperkok, C.C. Friedel, R. Zimmer, R. Roberts, R. Baric, J. Haas, Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome. PLoS ONE **2**, e459 (2007). doi:10.1371/journal.pone.0000459
54. K. Xu, B.J. Zheng, R. Zeng, W. Lu, Y.P. Lin, L. Xue, L. Li, L.L. Yang, C. Xu, J. Dai, F. Wang, Q. Li, Q.X. Dong, R.F. Yang, J.R. Wu, B. Sun, Severe acute respiratory syndrome coronavirus accessory protein 9b is a virion-associated protein. Virology **388**, 279–285 (2009)
55. W.S. Chan, C. Wu, S.C. Chow, T. Cheung, K.F. To, W.K. Leung, P.K. Chan, K.C. Lee, H.K. Ng, D.M. Au, A.W. Lo, Coronaviral hypothetical and structural proteins were found in the intestinal surface enterocytes and pneumocytes of severe acute respiratory syndrome (SARS). Mod. Pathol. **18**, 1432–1439 (2005)
56. M. Qiu, Y. Shi, Z. Guo, Z. Chen, R. He, R. Chen, D. Zhou, E. Dai, X. Wang, B. Si, Y. Song, J. Li, L. Yang, J. Wang, H. Wang, X. Pang, J. Zhai, Z. Du, Y. Liu, Y. Zhang, L. Li, J. Wang, B. Sun, R. Yang, Antibody responses to individual proteins of SARS coronavirus and their neutralization activities. Microbes Infect. **7**, 882–889 (2005)
57. Chinese SARS Molecular Epidemiology Consortium, Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science **303**, 1666–1669 (2004)
58. K.S. Saikatendu, J.S. Joseph, V. Subramanian, B.W. Neuman, M.J. Buchmeier, R.C. Stevens, P. Kuhn, Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. J. Virol. **81**, 3913–3921 (2007)
59. C. Meier, A.R. Aricescu, D.I. Stuart, J. Grimes, R.J.C. Gilbert, R.T. Aplin, R. Assenberg, The crystal structure of ORF 9b, a lipid binding protein from the SARS Coronavirus. Structure **14**, 1157–1165 (2006)
60. P. Stothard, The sequence manipulation suite, JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques **28**, 1102–1104 (2000)
61. P.M. Sharp, W.H. Li, The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15**, 1281–1295 (1987)
62. A. Pavesi, B. De Iaco, M.I. Granero, A. Porati, On the informational content of overlapping genes in prokaryotic and eukaryotic viruses. J. Mol. Evol. **44**, 625–631 (1997)
63. A. Pavesi, G. Magiorkinis, D.G. Karlin, Viral proteins originated de novo by overprinting can be identified by codon usage, application to the "gene nursery" of deltaretroviruses. PLoS Comput. Biol. **9**, e1003162 (2013). doi:10.1371/journal.pcbi.1003162

64. M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0. Bioinformatics **23**, 2947–2948 (2007)
65. P. Librado, J. Rozas, DnaSP v5, A software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25**, 1451–1452 (2009)
66. T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41**, 95–98 (1999)
67. Z. Obradovic, K. Peng, S. Vucetic, P. Radivojac, C.J. Brown, A.K. Dunker, Predicting intrinsic disorder from amino acid sequence. Proteins **53**, 566–572 (2003)
68. L.D. Hurst, The Ka/Ks ratio, diagnosing the form of sequence evolution. Trends Genet. **8**, 486 (2002)
69. W. Lapps, B.G. Hogue, D.A. Brian, Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. Virology **157**, 47–57 (1987)
70. F. Fischer, D. Peng, S.T. Hingley, S.R. Weiss, P.S. Masters, The internal open reading frame within the nucleocapsid gene of mouse hepatitis virus encodes a structural protein that is not essential for viral replication. J. Virol. **71**, 996–1003 (1997)
71. S.D. Senanayake, D.A. Brian, Bovine coronavirus I protein synthesis follows ribosomal scanning on the bicistronic N mRNA. Virus Res. **48**, 101–105 (1997)
72. E. Calvo, M.L. DeDiego, P. Garcia, J.A. Lopez, P. Perez-Brena, A. Falcon, Severe acute respiratory syndrome coronavirus accessory proteins 6 and 9b interact in vivo. Virus Res. **169**, 282–288 (2012)
73. C. Chang, M.H. Hou, C.F. Chang, C.D. Hsiao, T.H. Huang, The SARS coronavirus nucleocapsid protein - forms and functions. Antiviral Res. **103**, 39–50 (2014)